Annex VII

Evaluation of the Robustness of the SI Cutoff Criteria Used for the LLNA: BrdU-ELISA and LLNA: DA Test Methods

This page intentionally left blank

1.0 Evaluation of the Robustness of the SI Cutoff Criteria Used for the LLNA: BrdU-ELISA and LLNA: DA Test Methods

The analyses described in this annex aim to determine the robustness of the optimum SI criteria for the LLNA: BrdU-ELISA and LLNA: DA test methods. The analyses show that the optimal SI criteria for the LLNA: DA and the LLNA: BrdU-ELISA test methods are quite stable. Taking different samples of the data as training/validation sets has relatively little impact on the cutoff SI criteria or on the resulting number of false positives or false negatives. Both assays perform quite well for the optimized SI cutoff criteria. The proposed SI cutoff criteria should be adopted for now and re-optimized in the future after new prospective data have been collected.

1.1 Basis for Selection of the Optimized Criteria

The optimum SI criteria proposed in **Section 6.5** of the BRD were based on selecting the highest SI values that produced no false negatives, relative to traditional LLNA outcomes, in the entire databases of 43 (LLNA: BrdU-ELISA) or 44 (LLNA: DA) substances. Substances with multiple test results are represented by the most prevalent outcome for the SI criterion evaluated (e.g., if a substance had more negative than positive results at SI $\geq$ 1.6, then the substance was deemed negative). If there were an equal number of positive and negative tests for a substance at a particular SI criterion, then a conservative approach was taken where the substance was deemed positive at that criterion in order to be protective of public health. The "most prevalent outcome" approach is the same as using the median SI, or the higher of the two SI values in the middle of the data if there are an equal number of SI values.

1.2 Methods

Since there are no newly tested substances for which the optimized cutoff criteria (currently proposed to be SI $\geq$ 1.6 for the LLNA: BrdU-ELISA test method and SI $\geq$ 1.8 for the LLNA: DA test method) could be prospectively applied, a retrospective evaluation was performed. This retrospective validation involved taking various samples of the existing data as training sets, re-optimizing the SI cutoff criteria, and then applying the new criteria to the remainder of the data, which would serve as a validation set.

Such a validation exercise can be useful for situations in which the decision criteria for distinguishing between "positives" and "negatives" are quite complex and involve multiple variables. In such cases, it is quite common to discover that an apparently "successful" decision criteria based on a training set is really just an artifact unique to those substances, and cannot be generalized or extrapolated to another set of substances, such as a validation set. However, the LLNA: BrdU-ELISA and LLNA: DA criteria are extremely simple – a single SI cutoff value, which nevertheless produces an outstanding performance: no false negatives and only two false positives (<5%) for 43 LLNA: BrdU-ELISA-tested substances, and no false negatives and only three false positives (<7%) for the 44 LLNA: DA-tested substances. This excellent performance for a single SI cutoff criterion strongly argues that the criterion is robust to sampling.

When carrying out a validation exercise for the LLNA: BrdU-ELISA and LLNA: DA data, it is important to understand that only a small number of substances actually contribute to the determination and stability of the SI cutoff criterion. Thus, rather than taking various samples of the total dataset, one possible approach is a complete enumeration of all possible samples as it relates to the critical substances. Thus, one validation exercise carried out for the LLNA: BrdU-ELISA and LLNA: DA datasets was to look at all possible sample combinations of the four critical substances and examine the robustness of the optimized cutoff criterion in each case. In addition, a more traditional validation exercise for both the LLNA: BrdU-ELISA and LLNA: DA datasets was

performed. The datasets were first divided into phase I and phase II groups based on the dates that the data were submitted to NICEATM. The phase I substances were considered to be the training set, and the phase II substances were considered to be the validation set (and vice versa).

1.3 LLNA: BrdU-ELISA Results

The LLNA: BrdU-ELISA data for 43 substances are summarized and organized by test phase in **Table C-VII-1**. The decision rule applied to the data and the corresponding SI cutoff point were designed to minimize false positives while eliminating false negatives. As indicated above, the results were impressive, with a very low (<5%) false positive rate, when using SI $\geq$ 1.6 as the cutoff point.

It was noted that choosing SI $\geq$ 1.5 would produce exactly the same result as SI $\geq$ 1.6 for the 43 LLNA: BrdU-ELISA substances (no false negatives; two false positives). Choosing the lower critical value of 1.5 would minimize the likelihood of a false negative in the testing of future substances, while SI $\geq$ 1.6 minimizes the likelihood of future false positives. The calculations that follow use SI $\geq$ 1.6 as the critical cutoff. This same issue arises for the LLNA: DA data (see **Section 1.4** of this annex**)**. The SI $\geq$ 1.6 criterion was selected for the LLNA: BrdU-ELISA database because it was the highest SI value that produced no false negatives with minimal false positives.

For the first analysis, half of the LLNA: BrdU-ELISA substances were sampled to form a training set, while the remainder of the data served as the validation set. For each sample, the SI cutoff was re-optimized using the substances in the training set and then applied to the validation set. Because the criterion must be optimized to prevent false negatives and minimize the number of false positives, the SI cutoff is determined solely by the smallest positive SI response of the true positive substances in the training set. Thus, in a sample, the cutoff SI can only increase, never decrease, relative to the cutoff SI for entire database. Similarly, the false positive rate in the validation set can only go down, while the false negative rate can and does go up based on the cutoff value selected using the training set.

Table C-VII-1 SI Data for the LLNA: BrdU-ELISA[1]

Substance Name	SI for True Positives[2]	Substance Name	SI for True Negatives[3]
Phase I (N=31)			
Citral	16.35	Hexane	1.89
1, 4-Phenylenediamine	14.70	Lactic acid	1.89
Glutaraldehyde	14.60	Methyl salicylate	1.43
Diphenylcyclopropenone	11.62	Glycerol	1.29
Trimellitic anhydride	7.85	Dimethyl isophthalate	1.26
p-Benzoquinone	6.90	Propylene glycol	1.20
2, 4-Dinitrochloro-benzene	6.84	2-Hydroxypropyl-methacrylate	1.13
Isoeugenol	6.73	Isopropanol	1.01
Cyclamen aldehyde	5.71	Diethyl phthalate	0.88
Hydroxycitronellal	4.78		
Linalool	4.65		
Formaldehyde	4.40		
Isopropyl myristate	4.19		
Cinnamic aldehyde	3.97		
trans-Cinnamaldehyde	3.50		

continued

Table C-VII-1 SI Data for the LLNA: BrdU-ELISA[1] (continued)

Substance Name	SI for True Positives[2]	Substance Name	SI for True Negatives[3]
Phase I (N=31)			
Hexyl cinnamic aldehyde	3.40		
Eugenol	3.30		
3-Aminophenol	3.06		
Nickel sulfate	2.66		
4-Chloroaniline	2.53		
Aniline	2.07		
2-Mercaptobenzo-thiazole	1.62		
Phase II (N=12)			
Diethyl maleate	6.27	Salicylic acid	1.26
Ethyl acrylate	4.95	Sulfanilamide	1.26
5-Chloro-2-methyl-4-isothiazolin-3-one solution	4.83		
4-Methylaminophenol sulfate	3.98		
Cobalt chloride	3.68		
Phenyl benzoate	3.37		
Ethylene glycol dimethacrylate	3.11		
Cinnamic alcohol	2.74		
Sodium lauryl sulfate	2.64		
Imidazolidinyl urea	1.61		

Abbreviations: N = number of substances; SI = stimulation index.

[1] Substances with multiple test results are represented by the median SI, or the highest of the two SI values in the middle of the data if there are an equal number of SI values.

[2] True positive are substances that are positive in the traditional LLNA.

[3] True negatives are substances that are negative in the traditional LLNA.

The most critical substances for the LLNA: BrdU-ELISA data when evaluating the stability of the cutoff SI are the four lowest SI values for traditional LLNA positive substances. All of the 16 possible combinations of these substances are provided in **Table C-VII-2**.

Table C-VII-2 All Possible Distributions of Four Key Substances in Training (T) or Validation (V) Sets for LLNA: BrdU-ELISA

4-Chloro-aniline (SI=2.53)	Aniline (SI=2.07)	2-Mercapto-benzothiazole (SI=1.62)	Imidizolidinyl urea (SI=1.61)	Cutoff SI[1]	Validation Set	
					No. False Positives[2]	No. False Negatives
T	T	T	T	1.6	0-2	0
T	T	T	V	1.6	0-2	0
T	T	V	T	1.6	0-2	0
T	T	V	V	2.0	0	2
T	V	T	T	1.6	0-2	0
T	V	T	V	1.6	0-2	0
T	V	V	T	1.6	0-2	0
T	V	V	V	2.5	0	3
V	T	T	T	1.6	0-2	0
V	T	T	V	1.6	0-2	0
V	T	V	T	1.6	0-2	0
V	T	V	V	2.0	0	2
V	V	T	T	1.6	0-2	0
V	V	T	V	1.6	0-2	0
V	V	V	T	1.6	0-2	0
V	V	V	V	>2.5	0	≥4

Abbreviations: No. = number; SI = stimulation index; T= substance was in the training set; V = substance was in the validation set.

[1] The cutoff value is determined using the training set.

[2] The number of false positives in the validation set depend upon whether the two LLNA: BrdU-ELISA false positives with SI > 1.6, lactic acid (SI = 1.89) and hexane (SI = 1.89), are in the training set or in the validation set.

The cutoff SI values are relatively stable for the LLNA: BrdU-ELISA. The likelihood is 75% (12/16) that a validation exercise would result in an unchanged cutoff of SI ≥ 1.6, which also was the case when the phase I substances were used as the training set and the phase II substances were used as the validation set (and vice versa). The likelihood is 12.5% (2/16) that the cutoff will be elevated to SI ≥ 2, 6.25% (1/16) that it will be elevated to SI ≥ 2.5, and also 6.25% (1/16) that the re-optimized cutoff SI will exceed 2.5. The higher the cutoff SI, the greater the number of false negatives, as can be seen from **Table C-VII-2**. It is also important to recognize that most of the data are not relevant to determining the cutoff SI point. Only the "weakest positives" are critical, and the greater the variability among the SI values for these critical substances, the less stable the cutoff SI points will be.

The second validation exercise considered the phase I substances as a training set and the phase II substances as a validation set (and vice versa). If the phase I data are used as the training set, the SI cutoff point remains unchanged at ≥1.6; if the phase II data are used as the training set, then the SI cutoff point also remains unchanged (≥1.6). If the phase I data cutoff point was used in the evaluation of phase II substances, then there would be no false positives or false negatives. Conversely, if the phase II cutoff point was used to evaluate the substances in phase I, then there would be no false negatives and two false positives. Once again, the results of the validation study produce quite stable results.

1.4 LLNA: DA Results

The LLNA: DA data for 44 substances are organized by test phase and summarized in **Table C-VII-3**. Again, the decision rule applied to the data and the corresponding SI cutoff point were designed to minimize false positives while totally eliminating false negatives. These data showed a low (<7%) false positive rate. The cutoff value was set at SI $\geq$ 1.8 based on the data from the 44 substances, although a lower cutoff point, namely SI $\geq$ 1.7, would have performed exactly the same for these 44 substances (no false negatives; three false positives).

For the first analysis, half of the LLNA: DA substances were sampled to form a training set, while the remainder of the data served as a validation set. For each sample, the SI cutoff is re-optimized based on the substances in the training set and then applied to the validation set. Because the criterion must be optimized to prevent false negatives and minimize the number of false positives, the SI cutoff is determined solely by the smallest SI responses of the true positive substances in the training set. Thus in a sample, the cutoff SI can only increase, never decrease, relative to the cutoff SI for entire database. Similarly, the false positive rate in the validation set can only go down, while the false negative rate can and does go up based on the cutoff value selected using the training set.

Table C-VII-3 SI Data for the LLNA: DA[1]

Phase I (N=31)			
Substance Name	**SI for True Positives[2]**	**Substance Name**	**SI for True Negatives[3]**
2, 4-Dinitrochloro-benzene	9.96	Chlorobenzene	2.44
Isoeugenol	7.09	Hexane	2.31
Eugenol	7.07	1-Bromobutane	1.65
Benzalkonium chloride	6.68	Methyl salicylate	1.55
Abietic acid	6.26	Propylparaben	1.28
Hydroxycitronellal	5.69	Dimethyl isophthalate	1.26
Hexyl cinnamic aldehyde	5.50	Isopropanol	1.21
Phthalic anhydride	5.49	Diethyl phthalate	1.09
Potassium dichromate	5.49	Lactic acid	0.97
p-Phenylenediamine	5.14		
Glutaraldehyde	5.00		
Trimellitic anhydride	4.96		
Formaldehyde	4.84		
Cinnamic aldehyde	4.73		
Imidazolidinyl urea	4.67		
Citral	4.40		
Resorcinol	4.33		
Cobalt chloride	4.25		
Sodium lauryl sulfate	3.39		
3-Aminophenol	2.38		
Nickel (II) sulfate hexahydrate	2.13		
2-Mercaptobenzo-thiazole	2.00		

continued

Table C-VII-3 SI Data for the LLNA: DA[1] (continued)

Phase II (N=13)

Substance Name	SI for True Positives[2]	Substance Name	SI for True Negatives[3]
5-Chloro-2-methyl-4-isothiazolin-3-one	7.50	Salicylic acid	2.00
Cinnamic alcohol	5.66	Nickel (II) chloride	1.30
Propyl gallate	4.95	Sulfanilamide	0.86
Butyl glycidyl ether	4.59		
Ethylene glycol dimethacrylate	4.45		
Ethyl acrylate	4.29		
Phenyl benzoate	4.24		
p-Benzoquinone	3.79		
Diethyl maleate	3.78		
Methyl methacrylate	1.81		

Abbreviations: N = number of substances; SI = stimulation index.

[1] Substances with multiple test results are represented by the median SI or the highest of the two SI values in the middle of the data if there are an equal number of SI values.

[2] True positives are substances that are positive in the traditional LLNA.

[3] True negatives are substances that are negative in the traditional LLNA.

The four most critical substances for the LLNA: DA data when evaluating the stability of the cutoff SI are the four lowest SI values for positive substances. All of the 16 possible combinations of these substances are given in **Table C-VII-4**.

Table C-VII-4 All Possible Distributions of Four Key Substances in Training (T) or Validation (V) Sets for LLNA: DA

3-Aminophenol (SI=2.38)	Nickel sulfate (SI=2.13)	2-Mercapto-benzothiazole (SI=2.00)	Methyl methacrylate (SI=1.81)	Cutoff SI[1]	Validation Set	
					No. False Positives[2]	No. False Negatives
T	T	T	T	1.8	0-3	0
T	T	T	V	2.0	0-3	1
T	T	V	T	1.8	0-3	0
T	T	V	V	2.1	0-2	2
T	V	T	T	1.8	0-3	0
T	V	T	V	2.0	0-3	1
T	V	V	T	1.8	0-3	0
T	V	V	V	2.3	0-2	3
V	T	T	T	1.8	0-3	0
V	T	T	V	2.0	0-3	1
V	T	V	T	1.8	0-3	0
V	T	V	V	2.1	0-2	2

continued

Table C-VII-4 All Possible Distributions of Four Key Substances in Training (T) or Validation (V) Sets for LLNA: DA (continued)

3-Aminophenol (SI=2.38)	Nickel sulfate (SI=2.13)	2-Mercapto-benzothiazole (SI=2.00)	Methyl methacrylate (SI=1.81)	Cutoff SI[1]	Validation Set	
					No. False Positives[2]	No. False Negatives
V	V	T	T	1.8	0-3	0
V	V	T	V	2.0	0-3	1
V	V	V	T	1.8	0-3	0
V	V	V	V	>2.3	0-2	≥4

Abbreviations: No. = number; SI = stimulation index; T= substance was in the training set; V = substance was in the validation set.

[1] The cutoff value is determined using the training set.

[2] The number of false positives in the validation set depends upon whether the three LLNA: DA false positives (salicylic acid [SI = 2.0], hexane [SI= 2.31], and chlorobenzene [SI = 2.44]) are in the training set or in the validation set.

The cutoff SI values are relatively robust for the LLNA: DA test method also. The likelihood is 50% (8/16) that a validation exercise would result in an unchanged cutoff of SI ≥ 1.8. The likelihood is 25% (4/16) that the cutoff will be increased slightly to SI ≥ 2.0. The likelihood is 12.5% (2/16) that the cutoff will be elevated to SI ≥ 2.1 and 6.25% (1/16) that it will be greater than 2.3.

This conclusion regarding the stability of the cutoff SI is supported by the phase I vs. phase II approach to validation. This approach considered the phase I substances as a training set and the phase II substances as a validation set (and vice versa). If the phase I LLNA: DA data are used as the training set, the optimized cutoff SI criterion increases slightly from 1.8 to 2.0. If the phase II data are used as the training set, then the SI cutoff criterion remains unchanged at ≥1.8. If the phase I data cutoff point was used in the evaluation of phase II substances, then there would be one false positive and one false negative (methyl methacrylate, SI ≥ 1.81). Conversely, if the phase II cutoff point was used to evaluate the substances in phase I, then there would be no false negatives and two false positives.

1.5 Conclusions

These analyses show that the SI criteria for the LLNA: DA and LLNA: BrdU-ELISA test methods are quite robust. Taking different samples of the data as training/validation sets has relatively little impact on cutoff SI criteria or on the number of false positives or false negatives. Both assays perform quite well for the optimized SI cutoff criteria. The proposed SI cutoff criteria should be adopted for now, and re-optimized in the future after new prospective data have been collected.